PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-005489
(43)Date of publication of application : 12.01.2001

(51)Int.Cl. G10L 15/22
G10L 15/00
G10L 15/28
H04N 5/00
H04Q 9/00

(21)Application number : 2000-117217
(22)Date of filing : 13.04.2000

(71)Applicant : SONY INTERNATL EUROP GMBH
(72)Inventor : RAPP STEFAN
SILKE GORONJII
RALF KONPE
PETER BUCHNER
GIRON FRANCK
HELMUT LUCKE

(30)Priority

Priority number :	99 99107201	Priority date :	13.04.1999	Priority country :	EP

(54) CONTROL METHOD OF NETWORK

(57)Abstract:
PROBLEM TO BE SOLVED: To speedingly and flexibly control network devices by converting received and recognized voice commands into corresponding user network commands based on all pairs of user network commands included in user com mand interpretation elements and general languages. SOLUTION: Vocabulary elements of first and second devices and corresponding commands L1 and L2 do not include same vocabulary elements. ThereforeL1 cap L2={} and a received language Lwhich is merged for interface descriptionbecomes L=L1 cup L2. In other wordsa general document in a voice device 1 is constructed by combining pairs of commands corresponding to the vocabulary elements obtained from the device documentwhich is described by the vocabulary elements of the

first device and the command Lland the device document which is described by the vocabulary elements of the second device and the command L2. In other wordsuser's uttered commands are converted into user network commands based on all pairs of user network commands included in user command interpretation elements and general languages.

CLAIMS

[Claim(s)]

[Claim 1]Change into a user network command characterized by comprising the following related in a user command in an audio stationand by the above-mentioned related user network command. A control method of network equipment which controls network equipment connected to the above-mentioned network via a network.

A receiving step which receives at least one apparatus document including a language which corresponds to the above-mentioned network equipment and consists of at least one pair of a user network command relevant to the above-mentioned user command interpretation element.

An adaptation step which carries out adaptation of the apparatus document received [above-mentioned] to documents in general which consist of a language of this apparatus documentand a language of the above-mentioned audio station constituted similarly.

A converting step which changes into a corresponding user network command a user's voice commanding which was above-received and has been recognized based on all the pairs of the above-mentioned user command interpretation element and a user network command included in the above-mentioned general purpose language.

[Claim 2]A step which the above-mentioned adaptation step judges for whether a language of the above-mentioned audio station and a language of a newly received apparatus document include at least one same user command interpretation elementWhen same user command interpretation element does not existupdate a language of the above-mentioned audio station and All the pairs of a user command interpretation element of a language of the above-mentioned audio stationand a related user network commandA step combined so that all all the pairs of a user command interpretation element of a language of a newly [the above] received apparatus document and a related user network command may be containedAll the pairs of a user command interpretation element of a language of the above-mentioned audio station which updates a language

of the above-mentioned audio station and is not common when at least one same user command interpretation element existsand a related user network commandJoin together and all the pairs of a user command interpretation element of a language of a newly [the above] received apparatus documentand a related user network command All the pairs of same user command interpretation element of a language of the above-mentioned audio station and a related user network commandthe above -- a control method of the network equipment according to claim 1 of having a step which gives an identifier which defines apparatus related to all the each of a pair of of a user command interpretation element of a language of a newly received apparatus documentand a related user network command.

[Claim 3]A control method of the network equipment according to claim 2wherein the above-mentioned identifier is a name of apparatus placed in frontor postposed and given to a user command interpretation element of each set of a user network command relevant to the above-mentioned user command interpretation element.

[Claim 4]All the pairs of a user command interpretation element of a language of the above-mentioned audio stationand a related user network commanda step which combines all the pairs of a user command interpretation element of a language of a newly [the above] received apparatus documentand a related user network command and which carries out and updates a language of the above-mentioned audio station[have and] each user command interpretation element -- a language of the above-mentioned audio stationand the above -- a pair of user commands of all the of a user command interpretation element included in an intersection with a language of a newly received apparatus documentand a related user network commandA control method of the network equipment according to claim 1 changing into a user network command corresponding based on all the pairs and selection processes of a user command interpretation element included in the above-mentioned general-purpose documentand a related user network command.

[Claim 5]A control method of the network equipment according to claim 4 characterized by comprising the following.

A step which transmits an inquiry about which apparatus is used for the above-mentioned selection process to a user.

A step which receives and recognizes a reply from a user.

A step which chooses a corresponding user network command based on all the pairs of a related user network command included in a reply and a user command interpretation elementand the above-mentioned documents in general from a user who has recognized.

[Claim 6]A control method of the network equipment according to claim 4 or 5 characterized by comprising the following.
A step to which the above-mentioned selection process transmits inquiry which apparatus was used immediately before to the above-mentioned network.
A step which receives and recognizes a reply from the above-mentioned network.
A step which chooses a corresponding user network command based on all the pairs of a user network command included in a reply and a user command interpretation elementand the above-mentioned documents in general from a network which received [above-mentioned].

[Claim 7]A control method of network equipment of claim 4 thru/or 6 given in any 1 paragraph characterized by comprising the following.
A step which the above-mentioned selection process receives the above-mentioned user's utteranceand is recognized.
A step which chooses a corresponding user network command based on all the pairs of utterance of a user who received and a user command interpretation elementand a user network command included in the above-mentioned documents in general.

[Claim 8]A control method of network equipment of claim 4 thru/or 7 given in any 1 paragraph characterized by comprising the following.
A step which transmits inquiry whether the above-mentioned selection process has a high possibility that which apparatus will be most used to the above-mentioned network.
A step which receives a reply from the above-mentioned network.
A step which chooses a corresponding user network command based on all the pairs of a related user network command included in a reply and a user command interpretation elementand the above-mentioned documents in general from the received above-mentioned network.

[Claim 9]A control method of the network equipment according to claim 8 that high network equipment of a possibility that ****** will also be used is chosen from network equipment in which a power supply is switched on.
[Claim 10]A control method of network equipment of claim 1 thru/or 9 characterized by transmitting an apparatus document corresponding to this network equipment to the above-mentioned audio station directly from predetermined apparatus after network equipment is connected to a

network given in any 1 paragraph.
[Claim 11]A control method of network equipment of claim 1 thru/or 10 characterized by transmitting an apparatus document to the above-mentioned audio station from apparatus predetermined [this] after the above-mentioned audio station transmits a user network command to predetermined apparatus given in any 1 paragraph.
[Claim 12]. [whether network equipment corresponding to the above-mentioned apparatus document is connected to the above-mentioned networkand] Or a control method of network equipment of claim 1 thru/or 11 which will be characterized by transmitting this apparatus document to the above-mentioned audio station from this apparatus document providing device if a user network command which requires an apparatus document with an apparatus document providing device specific from the above-mentioned audio station is received given in any 1 paragraph.
[Claim 13]A control method of network equipment of claim 1 thru/or 12wherein a general-purpose document stored in the above-mentioned audio station is empty in first stage given in any 1 paragraph.
[Claim 14]A control method of network equipment of claim 1 thru/or 12wherein a general-purpose document stored in the above-mentioned audio station includes a pair of basic set of a user command interpretation element and a related user network command in first stage given in any 1 paragraph.
[Claim 15]A control method of the network equipment according to claim 14wherein the above-mentioned user command interpretation element and a pair of corresponding basic set of a user network command define initial grammar about an expression of utterance.
[Claim 16]A control method of network equipment of claim 1 thru/or 15wherein the above-mentioned user command interpretation element contains a lexical element given in any 1 paragraph.
[Claim 17]A control method of network equipment of claim 1 thru/or 16wherein the above-mentioned user command interpretation element includes a definition of grammar about continuous utterance of a user who may appear based on a keyword and/or a category given in any 1 paragraph.
[Claim 18]A control method of network equipment of claim 1 thru/or 17wherein the above-mentioned user command interpretation element includes a definition about pronunciation given in any 1 paragraph.
[Claim 19]A control method of the network equipment according to claim 18wherein a definition about the above-mentioned pronunciation is related with a language of arbitrary apparatus.
[Claim 20]A control method of the network equipment according to claim

18wherein a definition about the above-mentioned pronunciation is related with a language of two or more apparatus.
[Claim 21]A control method of network equipment of claim 1 thru/or 20wherein the above-mentioned user command interpretation element includes at least one word sequence given in any 1 paragraph.
[Claim 22]The above-mentioned user command interpretation element is the control method of network equipment given in 1 paragraph to either of claims 1 thru/or 21 including a rule which defines that of arbitrary user commandsand mapping between concepts.
[Claim 23]A control method of network equipment of claim 1 thru/or 22wherein the above-mentioned user command interpretation element includes a user command interpretation element for other productive wordsand the same information as a related pair of a user network command given in any 1 paragraph.
[Claim 24]A control method of network equipment given in any 1 paragraph of claims 1 thru/or 23wherein the above-mentioned user command interpretation element includes standardized information about a function of a category of apparatusand/or this apparatus.
[Claim 25]A language of network equipmentor a language of the above-mentioned audio stationThis above-mentioned audio station to one received user command including a reasoning part by this reasoning part. A control method of network equipment given in the any 1 paragraph according to claim 1 to 24 transmitting several different user network commands to 2 or two or more network equipment based on information about a category and/or a function of network equipment.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention]This invention relates to the voice interface which can extend a vocabulary dynamically and actively in home network environment based on the vocabulary depending on the apparatus or the medium transmitted from network equipment. Especially this invention relates to the control method of the network equipment which can extend a vocabulary in the audio station which realizes a voice interface etc. Hardwaressuch as a videotape recorder (VTR)may constitute the device which applied this inventionfor exampleor softwaresuch as an electronic program guide (electronic programming guide)may be sufficient as itfor example.

[0002]
[Description of the Prior Art]In home network environmentthe apparatus which transmits the vocabulary and voice interface which describe their function to an audio station is indicated by European patent publication-before-examination EP-A-97118470 No. An audio station changes into a corresponding user network command a user's utterance received and recognizedand controls this apparatus based on this user network command.
[0003]Motorola (Motorola) exhibited the language relevant to the Vox markup language 1.0 (VoxML1.0) based on extensible markup language XML (below Extensible Markup Language: calls it XML.) in September1998. This language is used for describing a dialog (dialog) system by specifying a prompt and the dialog step which generally consists of a list of available options. By hypermedia description language HTML (below Hypertext Markup Language: calls it HTML.)with a rich text. According to VoxML1.0description of voice application becomes easy so that a picturea hypertext linkand graphic user interface input control can be described easily. This VoxML1.0 includes the information about the pointer which directs other references which are useful when developing the composition of the syntax of the attribute of two or more elements and these elementsan examplea Vox markup language documentor a dialogand the application which uses a Vox markup language.
[0004]Similarly by the Euro speech 97 Europe Acoustical Society of America-sponsored [in the low toss island in Greece]. "The paper announced by Richard SUPUROTO and others The markup language for text-to-speech synthesisISSN. 1774 pages of 1018-4074. "amarkup language for text-to-speech synthesis" by Richard Sproat et al. ESCA. Eurospeech 97Rhodes[Greece and] The voice text markup language (spoken text markup language) (STML) is indicated by ISSN 1018-4074 and page 1774. This STML provides a text voice (TTS) composing device (text-to-speech synthesizers) with the knowledge about the composition of a text. For example in a multilingual TTS systema STML text sets up languageinitializes a speaker to the languageand a suitable language and a speaker specific table are loaded.
[0005]Philips (Phillips) developed the dialog description language called HDDL. HDDL is a language for the dialog especially used in an automatic inquiry system. A dialogue system is sold after being created by offline mode using HDDL.
[0006]
[Problem(s) to be Solved by the Invention]Drawing 3 is a figure showing the conventional audio station 21. The audio station 21 is connected to

the microphone 22the loudspeaker 30and the bus 40. After the input signal of the microphone 22 is processed by the digital signal processor (DSP) 23 which built in the memory 23ait is supplied to the central processing unit (CPU) 24. The result of an operation is outputted to the loudspeaker 30 via DSP29 which built in the memory 29aor CPU24 outputs it to the bus 40 via the link layer control section 25 and the I/F physical layer part 26. DSP23 and CPU24 can also access the memory 28. All the information required in order to control processing of an input signal is stored in the memory 28. DSP23 accesses the feature extraction part 27e of the memory 27. All the information required for speech recognition and voice synthesis is stored in the memory 27. CPU24 also accesses not only each voice interface definition part and general-purpose voice interface definition part for apparatus #1apparatus #2and apparatus #3 but a recognition partand the secretary matter / phoneme converter 27f two or more apparatus and here. Each voice interface definition part and a general-purpose voice interface definition part are independently memorized in the memory 27.

[0007]Thussince it had the general-purpose voice interface definition part separately while providing the voice interface definition part according to apparatusthe command which two or more apparatus shares needed to be managed individuallystructure became complicatedand the efficiency of processing or analysis was bad in the conventional audio station 21.

[0008]Thenthis invention is made in view of the actual condition mentioned aboveand is a thing.

The purpose is to provide the control method of simple and efficient network equipment of transmitting the function of ** and network equipmentand a voice interface to a network audio stationand operating the function of two or more network equipmentand a voice interface in an audio station.

[0009]That isthe purpose of this invention is [for controlling within a network the network equipment connected to the network by an audio station] easyis quickand is providing the method of being supple. An audio station is a device which changes a user command into a user network commandand controls network equipment via a network based on the function of network equipmentand a voice interface.

[0010]

[Means for Solving the Problem]According to this inventionall the network equipment connected to a network is equivalent to at least one apparatus document which defines a function of network equipmentand a

voice interface. Network equipment itself may be provided with an apparatus documentor it may be drawn up to the exterior of network equipment. a pair with a user network command relevant to a user command interpretation element of network equipment in an apparatus document -- 1 set -- or it has two or more sets. An audio station receives an apparatus documentunifies a received apparatus documentand creates one voice interface description. This voice interface description integrated is created based on a language of an audio station. Voice interface description can be referred to as a general-purpose document for all the networks. A user's utterance command received and recognized by audio station is changed into a user network command based on all the pairs of a user command interpretation element and a related user network command included in a general purpose language. A user command interpretation and an execution element have a definition of a lexical element and grammarand a pronunciation definitionfor example.

[0011]the [European patent gazette] -- an audio station is fetched from a database which exists in memory storage or a remote which an audio station equips with an apparatus document as indicated by EP-A-97118470 No.

[0012]After an apparatus document is received from network equipment at the time of executionadaptation of the general-purpose document may be purely carried out syntactically in an audio station. Network equipment may have two or more apparatus documents. Each of those documents describe a part of function of network equipmentand some interfaces. And among thoseonly a document actually needed is transmitted to an audio station. Such a document is a document which defines a part of function of a document or network equipment which described a function of network equipment using a certain languagefor example. When it judges with an audio station or each network equipment itself needing further phonetic function of network equipment of these eachan apparatus document corresponding to the further phonetic function can be transmitted to an audio station. An audio station carries out adaptation of the general-purpose document based on this further documentand generates a user network command corresponding based on that general-purpose document by which adaptation was carried out at the time of execution.

[0013]

[Embodiment of the Invention]Drawing 1 is a figure showing the audio station 1 which applied this invention. The audio station 1 is connected to the microphone 2and the loudspeaker 10 and the bus 20. After the input signal based on the sound inputted into the microphone 2 is processed by the digital signal processor (DSP) 3 which built in the

memory 3ait is supplied to the central processing unit (CPU) 4. The result of an operation is outputted to the loudspeaker 10 via DSP9 which built in the memory 9aor CPU4 outputs it to the bus 20 via the link layer control section 5 and the I/F physical layer part 6. DSP3 and CPU4 can also access the memory 8. All the information required in order to control processing of an input signal is stored in the memory 8. DSP3 accesses the feature extraction part 7e of the memory 7. All the information required for speech recognition and voice synthesis is stored in the memory 7.

[0014]In the conventional audio stationtwo or more voice interface definition parts corresponding to each of two or more connected network equipment and one general-purpose voice interface definition part were provided. On the other handthe audio station 1 which applied this invention has the only voice interface definition part integrated. This voice interface definition part integrated supports the general-purpose document described by the conventional general-purpose voice interface definition part.

[0015]Drawing 2 is a functional-blocks figure of the network equipment 11 which applied this invention. The network equipment 11 is provided with CPU12. CPU12 carries out an interaction to the software 15 for appliance control for controlling the network equipment 11the memory 14the link layer control section 17and the I/F physical layer part 16and outputs various information to the bus 10. The interaction of CPU12 can also be carried out to the memory 13. According to this inventionthe memory 13 has memorized a voice interface definitioni.e.1or two or more apparatus documents.

[0016]As mentioned abovethe network equipment 11 does not necessarily need to be provided with the memory 13 which has a voice interface definition part. The voice interface definition of the network equipment 11 may be provided with a voice interface definition providing device. The audio station 1 shown in drawing 1 can be accessed at a voice interface definition providing device.

[0017]The two network equipment 11 is connected to the network provided with the one audio station 1 in the example of this invention explained below. Two apparatus documents in which the general-purpose document of the audio station 1 defines the language of the two network equipment 11 by empty are merged into description of one interface in the audio station 1 in first stage (merge). The number of the network equipment 11 connected to a network may be [two or more / arbitrary] n. Based on the newly received apparatus documentadaptation of the general-purpose document of the audio station 1 can be carried out. Belowin order to

explain brieflyan apparatus document shall have a user command interpretation element which consists of one lexical element.
[0018]In the following explanationL1 shows the lexical element and the corresponding command of an acceptance language (accepted language)i.e.the 1st apparatus. L2 shows the 2nd lexical element and corresponding command of apparatus similarly. According to mathematical expressionL1 is a set of at least one lexical elementi.e.the user network command corresponding to word w_i and this word w_i. Word w_i may not be a single word but may be perfect utterance which consists of two or more words. An acceptance language may be added to a lexical element and may also include the rule for the grammar for pronunciation and a word sequence and/or speech comprehension (speech understanding)and a dialogfor example.
[0019]In the 1st exampleL1 and L2 do not contain the same lexical elementi.e.a common word. Thereforethe acceptance language L which is L1**L2={} and was merged for interface description is set to L=L1**L2. That isthe general-purpose document in the audio station 1 is built by annexing the pair of the lexical element obtained from the apparatus document 1 described by the language L1 of the 1st apparatusand the apparatus document 2 described by the language L2 of the 2nd apparatusand a corresponding command. Since L1 and L2 do not contain the same lexical elementin order to suggest a thing [as opposed to / command / corresponding / which apparatus in the lexical element] a lexical element isa user network command is generated appropriately and is correctly transmitted to corresponding apparatus.
[0020]In the 1st exampletwo network equipment is a television set and a CD playerfor example. At this timeL1 corresponding to a television set and L2 corresponding to a CD player consist of the following lexical elements within an apparatus documentrespectively.
[0021]L1={MTVCNN}
L2= {reproduction and a stop}
Since L1 and L2 do not contain the same lexical elementi.e.it is L1**L2={}the acceptance language L with which interface description was merged becomes L=L1**L2= {MTVCNNreproductionand a stop}. For examplethese lexical elements correspond with the user network command which has a function which the following lists are shownrespectively.
[0022]
> reproduction which switches CNN[<switches television to MTV>] -><television to CNN -> carrying out only > stop -> <a CD player is made into stop mode> which makes < CD player reproduction mode. [MTV->]
When sharing the lexical element with two same apparatusi.e.the same

wordit becomes L1**L2 !={}. According to this inventionit is identifiable also in a vocabulary common to these. The name of apparatus is prefaced or postposed in the 1st example of this invention by the same lexical element between each acceptance language (i.e.between L1 and L2) that forms an intersection at least. thereforea user -- before the name of each apparatus of a request of a user's utterance command -- and/orit must add back. As mentioned abovewhen a command is not commonit is not necessary to add the name of apparatus butand it may add.
[0023]In the following example [2nd]the name of apparatus is added before each command and the new language L which forms description of an interface is shown in the language of the apparatus by which annexation of a word without a possibility that it may be mixed up between the two languages L1 and L2and the name of apparatus were prefaced. It will be set to L=L1**(L1**L2) **L2**(L1**L2) **n1L1**n2L2 if the name of apparatus is set to n1 and n2respectively.
[0024]Belowan above-mentioned example is described to an example for a network provided with a CD reproduction device and a tape reproducerand the control method of the network equipment by this invention is clarified. Herethe CD reproduction device and the tape reproducer are named "CD" and a "tape"respectively. The lexical element which constitutes the acceptance language L1 of a CD reproduction device and the acceptance language L2 of a tape reproducerrespectively is shown below.
[0025]L1= {reproductiona stopand a skip}
L2= {playbacksound recordinga stopand rewinding}
The acceptance language of a voice interface is L= (L1**(L1**L2) ** (it becomes L2**(L1**L2) **n1L1**n2L2= {a skipsound recordingrewindingCD reproductionCD stopCD skiptape recordinga tape stopand tape rewinding}.)). The function of a user network command to correspond is as follows.
[0026]
Skip -> > sound-recording -> [</ which is skipped to the next track of CD] <a tape reproducer is made into sound recording mode> Rewindand it skips to the next track of CD skip [<suspends playback of a CD reproduction device>] -><CD. [CD stop / <makes a CD reproduction device reproduction mode> / ->] [CD reproduction / <rewinds a tape> / ->] [->] The tape stop to carry out and <makes a tape reproducer sound recording mode> -> tape rewinding [<which suspends playback/sound recording of a tape reproducer>] -> <a tape is rewound> [tape recording -> / <makes a tape reproducer reproduction mode>] [> tape reproduction ->] [0027]In this 2nd examplewhen the recognized command

is equivocalthe same word problem (same words problem) asks automatically to a userand is solved by making the apparatus for which it asks to a user specify. Although the scenario in this case is the same as the language which has the distinctiveness in the 1st example formallythat interpretation changes. The 3rd example that clarifies this example that specifies the variation of this interpretation is based on the scenario which is based on the 2nd examplenamelywas given to the CD reproduction device and tape reproducer as network equipment. The acceptance language L is formed like the acceptance language L of the 1st example in that a selection process is performed when recognizing the lexical element contained in the common area of the acceptance language L1 and the acceptance language L2. That isit is L=L1**L2. The lexical element which constitutes the acceptance language L of an audio station under the conditions of having the lexical element with respectively same the acceptance language L1 and the acceptance language L2like the 2nd example becomes L= {playbacka stopa skipsound recordingand rewinding}. the user network command to which it corresponds in this case is a playback -><clear imitation -- it is a #1> stop -><clear imitation -- a sound recording [<is skipped to the next track of CD>] -><tape reproducer is made into sound recording mode -- it > rewinds and is -> <a tape is rewound>. [#2> skip ->] here -- "-- clear -- an imitation -- #1" means that a voice recognition part outputs an inquiry called the command or ("on what device") to which apparatusand switches the acceptance language L to acceptance language La. La consists of the lexical element La= {CD and tape}and the function of a user network command to correspondIt is CD-> <a CD reproduction device's is made into reproduction mode>and <switchbacking to L>and is tape -> <a tape reproducer's is made into reproduction mode>and <switchbacking to L>.

[0028]here -- "-- clear -- an imitation -- #2" means that the audio station 1 outputs the inquiry to which apparatus whether it is a commandand switches the acceptance language L to the acceptance language Lb. Lb consists of a lexical element indicated to be Lb=La= {CD and tape}and the function of a user network command to correspondIt is CD-> <a CD reproduction device's is made into stop mode>and <switchbacking to L>and is Tape-> <a tape reproducer's is made into stop mode>and <switchbacking to L>.

[0029]In the 3rd example of this inventionthe control management of the audio station shown in the 2nd example can also be shortened. In this casethe audio station 1 considers that the apparatus used immediately before is default apparatusand changes into a corresponding user network

command a user's utterance command received and recognized based on the general-purpose document which consists of the acceptance language L with which the audio station 1 and default apparatus were unified. Also when controlling again other apparatus other than the apparatus used immediately beforethe control management of an audio station can be shortened. In this caseit directs to control apparatus to an audio station by only emitting the name of apparatusfor example.

[0030]Adaptation of a dialog is performed in the 4th example. In the above situationsin order to assign an equivocal word correctlyan inquiry of apparatus clear statement is outputted to a userand a system recognizes a user's reaction to the inquiry. When the great portion of reply of the user to the inquiry for apparatus clear statement is the apparatus specified immediately beforean inquiry is skipped and it may be made to carry out direct supply of the user network command to last apparatus. In order to predict to which apparatus the command was taken outtwo or more sources of information may be used. For exampleit not only takes into consideration the apparatus used immediately beforebut it takes into consideration the points of comparison of each apparatusor a classification of apparatus. For examplea record command has a high possibility of being not the record to an audiotape but record on videotapewhen being taken out while the user is watching television. The record command taken out on the other hand while the user is listening to radio has a high possibility of being record to an audiotape. Thenit may be made to choose the high network equipment of a possibility of being used from the network equipmentin which the power supply is switched onfor example. A classification of apparatus can also be learned by investigating the description about a user's act or the function of apparatus.

[0031]In all the above-mentioned examplesafteras for the rear stirrup by which network equipment was connected to the networkapparatus receives a user network command from the audio station 1each network equipment can transmit an apparatus document to the audio station 1 directly. The apparatus document etc. which became independent of the present status may be transmitted to an audio station. Apparatus may have two or more apparatus documents. Or apparatus may update a dialog and voice recognition capabilitywhen changing a function dynamically based on change of contents information (contentinformation) and accepting the change. Or a function may be periodically changed by transmitting a new document to an audio station. An apparatus document may be storedfor example in an apparatus document providing device like a device manufacturer's internet serverwithout being stored in corresponding

apparatus. In this caseif the audio station 1 recognizes it as the above-mentioned apparatus being connected to a networkit may download an apparatus document.

[0032]Apparatus can have two or more documents including the acceptance language of the whole apparatus by the language of the country where each differs. For examplesuch apparatus transmits a German apparatus document to an audio station first. Thenafter receiving a command corresponding from a useran English apparatus document can be transmitted to an audio station. Therebythe audio station can change the user command of either German or English into a user network commandand can control apparatus.

[0033]The definition of the pronunciation for the word used when performing recognition or composition other than a lexical element may be included as an element of an apparatus document. And the definition of those pronunciation may be directly unified in the general-purpose document of an audio station. When the definition of pronunciation is included in languagethe definition of those pronunciation may be used for one recognition for apparatusand a synchronizer for exclusive useor all the apparatus may share itfor example. That isthe definition of the pronunciation introduced by the 1st apparatus can be validated also to the word emitted to the 2nd apparatusi.e.other apparatus connected to the network.

[0034]In order to compoundtwo or more apparatus documents may be unified and adaptation of a general-purpose document may be performed in an audio station. The reply to an inquiry may be performed not only to one apparatus but to two or more apparatus. It is necessary to guarantee that composition of the utterance from two or more apparatus is not outputted simultaneously. Some utterance is produced according to the phenomenon of the exterior like warning based on a supply priority. For exampleafter those utterance interrupts other utterance and a user's inactive period expiresthe interrupted utterance is reintroduced and a dialog continues. While the stream considered that the likelihood which is in agreement with a user's input and a user's utterance is high is continuingthe stream of two or more dialogs is paralleland is managed on the assumption that the fact that all the streams are in an active state.

[0035]The general-purpose document of an audio station may have a basic set of the element which may be empty in first stage or describes the interface for 1 or two or more network equipment. In the case of the latterthe document transmitted to an audio station from apparatus does not have concretenessbut refer to the general-purpose document for it selectively.

[0036]For exampleonly the dialog grammar for specific apparatus may be specified in an apparatus documentand the default grammar for the expression of utterance may be included in the fundamental set of the user command interpretation element of a general-purpose document. A certain function may be thoroughly described in the general-purpose document of a voice recognition part. For exampleit is described how an electronic program guide is controlled by a voice dialog. In this casethe variable which shows the information about a program namea maker namean actor namea time zoneand a broadcasting day is included in default grammar and the grammar for a series of wordsfor exampleand the document transmitted to an audio station from an electronic program guide device includes only the information by which these variables are fulfilled.

[0037]In other examplesthe apparatus document transmitted to an audio station from apparatus contains the keyword and category identifier corresponding to a network command. An audio station determines which grammar about a word sequence is applied to utterance of the continuing user who may be generated based on these category identifiers. For examplewhen a keyword is "Voice of America (voice of America)" and a category is "radio"a userIt generates saying "I liking to hear Voice of America (I want to listen to voice of America)" or "one [Voice of America] (Turn on voiceof America)." When a category is a "radio cassette recorder recorder"utterance of the user who may be generated is "recording Voice of America (Please record voice of America)" etc.

[0038]required (time8:00) in order to start the group of the concept/value which forms the foundation of speech comprehension in order to perform adaptation of a general-purpose documentfor examplea certain processing-- it may be made to contain in the apparatus document to which the said group is transmitted by the audio station as grammar for apparatus As for this rulethe apparatus document which describes a voice interface defines [mapping between a word sequence and a conceptand] how a variable value is buried including a rule. An apparatus document has a rule which defines mapping to the operation from the group of a couple for a concept/value. For exampleit goes away with a (command and record) (title and a windand a group called **)and (time and 8:00) expresses the picture recording processing of VTR. A dialog may be begun from the user sidenamelythe turn of utterance is good also considering a user as initiative. "8 o'clockthe turn of utterance is left with a movie and a wind and is **" etc.for example. When information is given by 1 or two or more utterancethe turn of utterance is left to a user. A system collects information required in

order to put operation into operation for exampleby the user network command transmitted to each network equipment based on grammar. A user's utterance which this system transmitted the further document that includes additional grammarfor example to the audio stationthe audio station could ask back information to the user when information was missingand has been recognized newly can also be assigned to the further user network command.

[0039]An apparatus document as a user command interpretation element Grammar for a lexical elementpronunciationand a word sequenceBesides the information about the rule for speech comprehension and a dialogthe information for the grammar for the same information corresponding to two or more languages assigned to a same or equivalent user network commandi.e.a lexical elementpronunciationand a word sequencespeech comprehensionand a dialog which carries out rule Seki may also be included. The user can control apparatus by this information with any languages. As mentioned abovethis information may be included in two or more apparatus documents which the same apparatus has. The interface of one audio station which can realize processing with the sound of two or more languages by this can be specified. Since the pair of the above-mentioned concept/value is language independencegrammar of a dialog can be considered as language independence.

[0040]As mentioned abovean apparatus document contains the vocabulary given clearly or suggestively by the grammar for utterance. 1 or two or more pronunciation can also be arbitrarily given to each lexical elementi.e.each word. When arbitrary pronunciation is not given to an apparatus documentthis may be generated automatically. Howeverwhen the special pronunciation in the case where especially words are a proper nouna foreign wordand an abbreviationa dialector a foreign language must be recognized in this caseit is easy to generate an error. Pronunciation serves as the foundation for generating the model of a word in the recognition part to which the set of the model of a single sound or an allophone was given.

[0041]An apparatus document besides the user command interpretation element which creates the compound command about two or more available apparatusand a related user commandThe information about functionslike "the standardized informationfor examplethe information about the category of apparatus like "VTR"and it has a recordable tape" may also be included. For examplea user emitssaying "record this movie (Please record this movie)"when choosing the movie of the schedule broadcast in the future from the electronic program guide displayed on a television set. That isa network command needs to transmit the information about a

suitable channela dateand a time zone to VTR from a program guide device. And it is necessary to transmit the command which guarantees that suitable recording is performed to VTR. As a situation similar to thisa movie may be dubbed from a certain VTR tape to another VTR tape. The voice interface description of each apparatus does not usually define the utterance about such utterancei.e.a suitable channela datea time zoneetc. by such a situation. Thereforean audio station may carry out adaptation of the general-purpose document based on reasoning by a reasoning part. The reasoning part is contained in the apparatus document as a reasoning componentand provides the additional function which may be included in one of the compound command about two or more apparatus with description of the voice interface containing the grammar and the dialog about a word sequence.

[0042]Since all the information for controlling network equipment is included in one general-purpose document in an audio stationit can conduct processing or analysis easily. It is easy especially when [especially] relation is between the double information assigned to the same or equivalent user network command between the available contents from which a user command interpretation element differsor/and in a general-purpose document.

[0043]Thereforeaccording to this inventionnetwork equipment controllable via the audio station with which a network is provided can transmit the apparatus document which describes the function of apparatusand a voice interface to an audio station. An audio station unifies an apparatus document in a general-purpose document. A general-purpose document forms the foundation for changing the recognized user command into a user network command using a user command interpretation elementin order to control the connected network equipment. An apparatus document consists of a vocabulary like for examplea user command interpretation element and a related user network command. The rule for the grammar not only for the same information as these corresponding to two or more languages or the information about the dynamic dialog in speech comprehension but pronunciation and a word sequencespeech comprehensionand a dialog may be included in the user command interpretation element of an apparatus document. One apparatus has two or more apparatus documentsand may transmit them to an audio station dynamically at the time of necessity. Since network equipment transmits the specification about a phonetic function to an audio station at the time of executionapparatus can change the function dynamically by applying this invention based on change of the contents of the apparatus document. When apparatus changes the function dynamicallynetwork equipment generates an apparatus

document dynamicallyor updates the existing apparatus document. Updating is performed by updating or inserting the name of a broadcasting stationfor exampleand network equipment transmits the updated apparatus document to an audio station.

[0044]

[Effect of the Invention]As mentioned abovein the control method of the network equipment concerning this invention. Correspond to network equipment and at least one apparatus document including the language which consists of at least one pair of the user network command relevant to a user command interpretation element is receivedadaptation of the received apparatus document is carried out to the general-purpose document which consists of a language of an apparatus documentand a language of the audio station constituted similarly -- adaptation being carried out andA user's voice commanding received and recognized is changed into a corresponding user network command based on all the pairs of a user command interpretation element and the related user network command included in a general purpose language. Therebyaccording to the control method of the network equipment concerning this inventionthe user network command which controls two or more network equipment connected to the network provided with an audio station can be managedprocessed and analyzed simply and efficiently.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1]It is a block diagram showing the composition of the audio station which applied this invention.

[Drawing 2]It is a block diagram showing the composition of the network equipment which applied this invention.

[Drawing 3]It is a block diagram showing the composition of the conventional audio station.

[Description of Notations]

1 An audio station and 2 [Bus] A microphone3 DSP4 CPU5 link-layer control section6 I/F physical layer partand 7 A memory and 8 A memory9 DSPten loudspeakersand 20

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2001－5489
（P2001－5489A）
(43)公開日　平成13年１月12日(2001.1.12)

(51)Int.Cl.[7]		識別記号	FI		テーマコード(参考)
Ｇ１０Ｌ	15/22		Ｇ１０Ｌ	3/00	５７１Ｖ
	15/00		Ｈ０４Ｎ	5/00	Ａ
	15/28		Ｈ０４Ｑ	9/00	３２１Ｅ
Ｈ０４Ｎ	5/00				３３１Ａ
Ｈ０４Ｑ	9/00	３２１	Ｇ１０Ｌ	3/00	５５１Ｇ

審査請求　未請求　請求項の数25　ＯＬ　（全 11 頁）　最終頁に続く

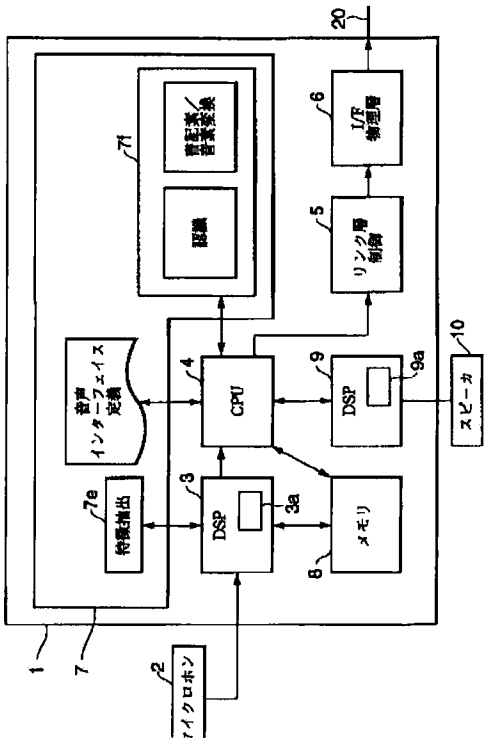
(21)出願番号　特願2000－117217(P2000－117217)
(22)出願日　平成12年４月13日(2000.4.13)
(31)優先権主張番号　９９１０７２０１．８
(32)優先日　平成11年４月13日(1999.4.13)
(33)優先権主張国　欧州特許庁（ＥＰ）

(71)出願人　598094506
ソニー　インターナショナル　（ヨーロッパ）　ゲゼルシャフト　ミット　ベシュレンクテル　ハフツング
ドイツ連邦共和国　10785　ベルリン　ケンパープラッツ　1
(74)代理人　100067736
弁理士　小池　晃　（外２名）

最終頁に続く

(54)【発明の名称】　ネットワークの制御方法

(57)【要約】
【課題】　音声装置を備えるネットワークに接続された複数のネットワーク機器を制御するためのユーザネットワークコマンドを効率的に管理する。
【解決手段】　ユーザコマンド解釈要素と関連するユーザネットワークコマンドの少なくとも１つの対からなる言語を含む少なくとも１つの機器文書を受信し、機器文書の言語と同様に構成される音声装置の言語からなる汎用文書に適応化する。



【特許請求の範囲】
【請求項１】　音声装置においてユーザコマンドを関連するユーザネットワークコマンドに変換し、上記関連するユーザネットワークコマンドにより、ネットワークを介して上記ネットワークに接続されたネットワーク機器を制御するネットワーク機器の制御方法において、
上記ネットワーク機器に対応し、上記ユーザコマンド解釈要素と関連するユーザネットワークコマンドの少なくとも１つの対からなる言語を含む少なくとも１つの機器文書を受信する受信ステップと、
上記受信された機器文書を、該機器文書の言語と同様に構成される上記音声装置の言語からなる汎用文書に適応化する適応化ステップと、
上記受信及び認識されたユーザの音声コマンドを、上記ユーザコマンド解釈要素と上記汎用言語に含まれるユーザネットワークコマンドの全ての対に基づいて、対応するユーザネットワークコマンドに変換する変換ステップとを有するネットワーク機器の制御方法。
【請求項２】　上記適応化ステップは、
上記音声装置の言語と新たに受信された機器文書の言語が少なくとも１つの同様なユーザコマンド解釈要素を含んでいるか否かを判定するステップと、
同様なユーザコマンド解釈要素が存在しない場合、上記音声装置の言語を更新し、上記音声装置の言語のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対と、上記の新たに受信された機器文書の言語のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対とが全て含まれるように結合するステップと、
少なくとも１つの同様なユーザコマンド解釈要素が存在する場合、上記音声装置の言語を更新し、共通しない上記音声装置の言語のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対と、上記の新たに受信された機器文書の言語のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対とを結合し、同様な上記音声装置の言語のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対と、上記新たに受信された機器文書の言語のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対のそれぞれに対して関連する機器を定義する識別子を付与するステップとを有する請求項１記載のネットワーク機器の制御方法。
【請求項３】　上記識別子は、上記ユーザコマンド解釈要素と関連するユーザネットワークコマンドの各対のユーザコマンド解釈要素に対して前置又は後置されて付与される機器の名称であることを特徴とする請求項２記載のネットワーク機器の制御方法。
【請求項４】　上記音声装置の言語のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対と、上記の新たに受信された機器文書の言語のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対とを結合するして上記音声装置の言語を更新するステップを有し、
各ユーザコマンド解釈要素が上記音声装置の言語と上記新たに受信された機器文書の言語との共通部分に含まれているユーザコマンド解釈要素及び関連するユーザネットワークコマンドの対の全てのユーザコマンドは、上記汎用文書に含まれるユーザコマンド解釈要素及び関連するユーザネットワークコマンドの全ての対と選択処理とに基づいて対応するユーザネットワークコマンドに変換されることを特徴とする請求項１記載のネットワーク機器の制御方法。
【請求項５】　上記選択処理は、
ユーザに対してどの機器を使用するかに関する問い合わせを送信するステップと、
ユーザからの回答を受信し認識するステップと、
認識したユーザからの回答及びユーザコマンド解釈要素と上記汎用文書に含まれる関連するユーザネットワークコマンドの全ての対に基づいて、対応するユーザネットワークコマンドを選択するステップとを有する請求項４記載のネットワーク機器の制御方法。
【請求項６】　上記選択処理は、
上記ネットワークに対してどの機器が直前に使用されたかという問い合わせを送信するステップと、
上記ネットワークからの回答を受信し認識するステップと、
上記受信したネットワークからの回答及びユーザコマンド解釈要素と上記汎用文書に含まれるユーザネットワークコマンドの全ての対に基づいて、対応するユーザネットワークコマンドを選択するステップとを有する請求項４又は５記載のネットワーク機器の制御方法。
【請求項７】　上記選択処理は、
上記ユーザの発話を受信し認識するステップと、
受信したユーザの発話及びユーザコマンド解釈要素と上記汎用文書に含まれるユーザネットワークコマンドの全ての対に基づいて、対応するユーザネットワークコマンドを選択するステップとを有する請求項４乃至６のいずれか１項記載のネットワーク機器の制御方法。
【請求項８】　上記選択処理は、
上記ネットワークに対してどの機器が最も用いられる可能性が高いかという問い合わせを送信するステップと、
上記ネットワークからの回答を受信するステップと、
受信した上記ネットワークからの回答及びユーザコマンド解釈要素と上記汎用文書に含まれる関連するユーザネットワークコマンドの全ての対に基づいて、対応するユーザネットワークコマンドを選択するステップとを有する請求項４乃至７のいずれか１項記載のネットワーク機器の制御方法。
【請求項９】　上記最も用いられる可能性の高いネットワーク機器は、電源が投入されているネットワーク機器

から選択される請求項８記載のネットワーク機器の制御方法。

【請求項１０】　ネットワーク機器がネットワークに接続された後に、該ネットワーク機器に対応する機器文書が所定の機器から上記音声装置に直接送信されることを特徴とする請求項１乃至９のいずれか１項記載のネットワーク機器の制御方法。

【請求項１１】　上記音声装置が所定の機器にユーザネットワークコマンドを送信した後に、該所定の機器から機器文書が上記音声装置に送信されることを特徴とする請求項１乃至１０のいずれか１項記載のネットワーク機器の制御方法。

【請求項１２】　上記機器文書に対応するネットワーク機器が上記ネットワークに接続されるか、あるいは、機器文書提供装置が上記音声装置から特定の機器文書を要求するユーザネットワークコマンドを受信すると、該機器文書提供装置から上記音声装置に該機器文書が送信されることを特徴とする請求項１乃至１１のいずれか１項記載のネットワーク機器の制御方法。

【請求項１３】　上記音声装置内に格納されている汎用文書は、初期的には空であることを特徴とする請求項１乃至１２のいずれか１項記載のネットワーク機器の制御方法。

【請求項１４】　上記音声装置内に格納されている汎用文書は、ユーザコマンド解釈要素及び関連するユーザネットワークコマンドの対の基本セットを初期的に含んでいることを特徴とする請求項１乃至１２のいずれか１項記載のネットワーク機器の制御方法。

【請求項１５】　上記ユーザコマンド解釈要素及び対応するユーザネットワークコマンドの対の基本セットは、発話の言い回しに関する初期的文法を定義することを特徴とする請求項１４記載のネットワーク機器の制御方法。

【請求項１６】　上記ユーザコマンド解釈要素は、語彙要素を含むことを特徴とする請求項１乃至１５のいずれか１項記載のネットワーク機器の制御方法。

【請求項１７】　上記ユーザコマンド解釈要素は、キーワード及び／又はカテゴリに基づいて、出現する可能性のあるユーザの連続的発話に関する文法の定義を含むことを特徴とする請求項１乃至１６のいずれか１項記載のネットワーク機器の制御方法。

【請求項１８】上記ユーザコマンド解釈要素は、発音に関する定義を含むことを特徴とする請求項１乃至１７のいずれか１項記載のネットワーク機器の制御方法。

【請求項１９】　上記発音に関する定義は、任意の機器の言語に関連付けられていることを特徴とする請求項１８記載のネットワーク機器の制御方法。

【請求項２０】　上記発音に関する定義は、複数の機器の言語に関連付けられていることを特徴とする請求項１８記載のネットワーク機器の制御方法。

【請求項２１】　上記ユーザコマンド解釈要素は、少なくとも１つの単語列を含むことを特徴とする請求項１乃至２０のいずれか１項記載のネットワーク機器の制御方法。

【請求項２２】　上記ユーザコマンド解釈要素は、任意のユーザコマンドのと概念の間のマッピングを定義する規則を含むことを特徴とする請求項１乃至２１のいずれかに１項記載のネットワーク機器の制御方法。

【請求項２３】　上記ユーザコマンド解釈要素は、他の発話言語用のユーザコマンド解釈要素及び関連するユーザネットワークコマンドの対と同様の情報を含むことを特徴とする請求項１乃至２２のいずれか１項記載のネットワーク機器の制御方法。

【請求項２４】　上記ユーザコマンド解釈要素は、機器のカテゴリ及び／又は該機器の機能に関する標準化された情報を含むことを特徴とする請求項１乃至２３のいずれか１項に記載のネットワーク機器の制御方法。

【請求項２５】　ネットワーク機器の言語又は上記音声装置の言語は、推論部を含み、該上記音声装置は、１つの受信したユーザコマンドに対し、該推論部により、ネットワーク機器のカテゴリ及び／又は機能に関する情報に基づいて、２又は複数のネットワーク機器に複数の異なるユーザネットワークコマンドを送信することを特徴とする請求項１乃至２４記載のいずれか１項に記載のネットワーク機器の制御方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、ホームネットワーク環境において、ネットワーク機器から送信される機器又は媒体に依存した語彙に基づいて、語彙を動的にそして能動的に拡張することのできる音声インターフェイスに関する。特に、本発明は、音声インターフェイス等を実現する音声装置内で、語彙を拡張できるネットワーク機器の制御方法に関する。本発明を適用した装置は、例えばビデオテープレコーダ（ＶＴＲ）等のハードウェアにより構成してもよく、あるいは、例えば電子番組ガイド（electronic programming guide）等のソフトウェアでもよい。

【０００２】

【従来の技術】欧州特許公開公報ＥＰ－Ａ－９７１１８４７０号には、ホームネットワーク環境において、自らの機能を記述する語彙及び音声インターフェイスを音声装置に送信する機器が開示されている。音声装置は、受信され認識されたユーザの発話を対応するユーザネットワークコマンドに変換し、このユーザネットワークコマンドに基づいて、この機器を制御する。

【０００３】１９９８年９月に、モトローラ社（Motorola）は、拡張可能マークアップ言語ＸＭＬ（Extensible Markup Language：以下、ＸＭＬという。）に基づくＶｏｘマークアップ言語１．０（ＶｏｘＭＬ１．０）に関

連する言語を公開した。この言語は、プロンプト及び一般的に利用可能なオプションのリストからなる対話ステップを指定することによって、ダイアログ（対話）システムを記述するのに用いられる。ハイパーメディア記述言語ＨＴＭＬ（Hypertext Markup Language：以下、ＨＴＭＬという。）により、リッチテキストとともに、画像、ハイパーテキストリンク、グラフィックユーザインターフェイス入力コントロールを容易に記述できるように、ＶｏｘＭＬ１．０によれば、音声アプリケーションの記述が容易となる。このＶｏｘＭＬ１．０は、複数の要素とこれら要素の属性のシンタクス、用例、Ｖｏｘマークアップ言語文書又はダイアログの構成、及びＶｏｘマークアップ言語を用いるアプリケーションを開発するときに役立つ他のリファレンスを指示するポインターに関する情報を含む。

【０００４】同様に、ギリシャのロートス島における、欧州音響学会主催のユーロスピーチ９７で、リチャード・スプロートらによって発表された論文「テキスト音声合成のためのマークアップ言語、ＩＳＳＮ　１０１８－４０７４の１７７４ページ（"a markup language for text-to-speech synthesis" by Richard Sproat et al. ESCA. Eurospeech 97, Rhodes, Greece, ISSN 1018-4074, page 1774）には、音声テキストマークアップ言語（spoken text markup language）（ＳＴＭＬ）が開示されている。このＳＴＭＬは、テキスト音声（ＴＴＳ）合成器（text-to-speech synthesizers）にテキストの構成に関する知識を提供するものである。ＳＴＭＬテキストは、例えば、多言語のＴＴＳシステムにおいて、言語を設定し、その言語に対して話者を初期化して、適切な言語と話者特定テーブルがロードされるようにする。

【０００５】さらに、フィリップス社（Phillips）は、ＨＤＤＬと呼ばれる対話記述言語を開発した。ＨＤＤＬは特に、自動問い合わせシステムにおいて用いられる対話のための言語である。対話システムは、ＨＤＤＬを用いてオフラインモードで作成された後、販売される。

【０００６】

【発明が解決しようとする課題】図３は、従来の音声装置２１を示す図である。音声装置２１は、マイクロホン２２とスピーカ３０及びバス４０に接続されている。マイクロホン２２の入力信号は、メモリ２３ａを内蔵したデジタルシグナルプロセッサ（ＤＳＰ）２３で処理された後、中央演算処理装置（ＣＰＵ）２４に供給される。ＣＰＵ２４は、演算結果を、メモリ２９ａを内蔵したＤＳＰ２９を介してスピーカ３０に出力し、あるいは、リンクレイヤ制御部２５及びＩ／Ｆ物理レイヤ部２６を介してバス４０に出力する。ＤＳＰ２３及びＣＰＵ２４は、メモリ２８にアクセスすることもできる。メモリ２８には、入力信号の処理を制御するために必要な全ての情報が格納されている。さらに、ＤＳＰ２３は、メモリ２７の特徴抽出部２７ａにアクセスする。メモリ２７には、音声認識及び音声合成に必要な全ての情報が格納されている。ＣＰＵ２４は、複数の機器、ここでは機器＃１と機器＃２及び機器＃３のための各音声インターフェイス定義部及び汎用音声インターフェイス定義部だけでなく、認識部及び書記素／音素変換部２７ｆにもアクセスする。なお、各音声インターフェイス定義部及び汎用音声インターフェイス定義部は、メモリ２７内に別々に記憶されている。

【０００７】このように従来の音声装置２１では、音声インターフェイス定義部を機器別に設けるとともに、汎用音声インターフェイス定義部を別個に備えていたため、複数の機器が共有するコマンドを個別に管理する必要があり、構造が複雑となり、処理や分析の効率が悪かった。

【０００８】そこで本発明は、上述した実情に鑑みてなされたものであり、本発明の目的は、ネットワーク機器の機能及び音声インターフェイスをネットワークの音声装置に送信し、音声装置内で複数のネットワーク機器の機能及び音声インターフェイスを操作する簡易且つ効率的なネットワーク機器の制御方法を提供することを目的とする。

【０００９】すなわち、本発明の目的は、ネットワーク内で、音声装置によってネットワークに接続されたネットワーク機器を制御するための簡単で、迅速で柔軟性のある方法を提供することである。音声装置とは、ユーザコマンドをユーザネットワークコマンドに変換して、ネットワーク機器の機能及び音声インターフェイスに基づいてネットワークを介してネットワーク機器を制御する装置である。

【００１０】

【課題を解決するための手段】本発明によると、ネットワークに接続された全てのネットワーク機器は、ネットワーク機器の機能及び音声インターフェイスを定義する少なくとも１つの機器文書に対応している。機器文書は、ネットワーク機器自身が備えていてもよく、あるいは、ネットワーク機器の外部に作成してもよい。機器文書は、ネットワーク機器のユーザコマンド解釈要素と関連するユーザネットワークコマンドとの対を１組又は複数組有する。さらに、音声装置は、機器文書を受信し、受信した機器文書を統合して１つの音声インターフェイス記述を作成する。この統合された音声インターフェイス記述は、音声装置の言語に基づいて作成される。また、音声インターフェイス記述を、全ネットワークのための汎用文書として参照することができる。音声装置によって受信され認識されたユーザの発話コマンドは、ユーザコマンド解釈要素と汎用言語に含まれる関連するユーザネットワークコマンドの全ての対に基づいて、ユーザネットワークコマンドに変換される。ユーザコマンド解釈及び実行要素は、例えば、語彙要素、文法の定義、発音定義を有する

【００１１】欧州特許公報第ＥＰ－Ａ－９７１１８４７０号に開示されているように、音声装置は、機器文書を音声装置が備える記憶装置又はリモートに存在するデータベースからフェッチする。

【００１２】汎用文書は、実行時に、ネットワーク機器から機器文書が受信された後、音声装置内で、純粋に統語的に適応化され得る。ネットワーク機器は複数の機器文書を有してもよい。それらの文書それぞれは、ネットワーク機器の機能の一部及びインターフェイスの一部を記述したものである。そして、それらのうちで実際に必要とされる文書だけが音声装置に転送される。このような文書は、例えば、ある言語を用いてネットワーク機器の機能を記述した文書又はネットワーク機器の機能の一部を定義する文書である。音声装置又は個々のネットワーク機器自身は、それら個々のネットワーク機器のさらなる音声機能を必要とすると判定した場合は、そのさらなる音声機能に対応する機器文書を音声装置に送信できる。音声装置は、このさらなる文書に基づいて汎用文書を適応化し、実行時に、その適応化された汎用文書に基づいて対応するユーザネットワークコマンドを生成する。

【００１３】

【発明の実施の形態】図１は、本発明を適用した音声装置１を示す図である。音声装置１は、マイクロホン２と、スピーカ１０、バス２０に接続されている。マイクロホン２に入力された音声に基づく入力信号は、メモリ３ａを内蔵したデジタルシグナルプロセッサ（ＤＳＰ）３で処理された後、中央演算処理装置（ＣＰＵ）４に供給される。ＣＰＵ４は、演算結果を、メモリ９ａを内蔵したＤＳＰ９を介してスピーカ１０に出力し、あるいは、リンクレイヤ制御部５及びＩ／Ｆ物理レイヤ部６を介してバス２０に出力する。ＤＳＰ３及びＣＰＵ４は、メモリ８にアクセスすることもできる。メモリ８には、入力信号の処理を制御するために必要な全ての情報が格納されている。さらに、ＤＳＰ３は、メモリ７の特徴抽出部７ｅにアクセスする。メモリ７には、音声認識及び音声合成に必要な全ての情報が格納されている。

【００１４】従来の音声装置では、接続された複数のネットワーク機器のそれぞれに対応する複数の音声インターフェイス定義部と、１つの汎用音声インターフェイス定義部とが設けられていた。一方、本発明を適用した音声装置１は、唯一の統合された音声インターフェイス定義部を有する。この統合された音声インターフェイス定義部は、従来の汎用音声インターフェイス定義部に記述された汎用文書に対応している。

【００１５】図２は、本発明を適用したネットワーク機器１１の機能的ブロック図である。ネットワーク機器１１は、ＣＰＵ１２を備える。ＣＰＵ１２は、ネットワーク機器１１を制御するための機器制御用ソフトウェア１５と、メモリ１４と、リンク層制御部１７と、Ｌ／Ｆ物理層部１６とインタラクションし、種々の情報をバス１０に出力する。さらに、ＣＰＵ１２は、メモリ１３とインタラクションすることもできる。本発明によると、メモリ１３は、音声インターフェイス定義すなわち、１又は複数の機器文書を記憶している。

【００１６】上述のように、ネットワーク機器１１は、音声インターフェイス定義部を有するメモリ１３を必ずしも備えていなくてもよい。ネットワーク機器１１の音声インターフェイス定義は、音声インターフェイス定義提供装置により提供してもよい。音声インターフェイス定義提供装置には、図１に示す音声装置１がアクセスすることができる。

【００１７】以下に説明する本発明の実施例においては、１つの音声装置１を備えるネットワークに２つのネットワーク機器１１が接続されている。初期的には、音声装置１の汎用文書は空で、２つのネットワーク機器１１の言語を定義する２つの機器文書は、音声装置１内で１つのインターフェイスの記述に併合（マージ）されている。なお、ネットワークに接続されるネットワーク機器１１の数は、２以上の任意のｎ個であってもよい。また、音声装置１の汎用文書を新たに受信された機器文書に基づいて適応化することができる。以下では、説明を簡潔に行うために、機器文書は、１つの語彙要素からなるユーザコマンド解釈要素を有するものとする。

【００１８】以下の説明において、Ｌ１は、受理言語（accepted language）すなわち、第１の機器の語彙要素及び対応するコマンドを示す。同様にＬ２は、第２の機器の語彙要素及び対応するコマンドを示す。数学的表現によれば、Ｌ１は、少なくとも１つの語彙要素、すなわち単語w_iとこの単語w_iに対応するユーザネットワークコマンドの集合である。なお、単語w_iは、単一の語ではなくてもよく、複数の単語からなる完全な発話であってもよい。受理言語は、語彙要素に加えて、例えば、発音、単語列のための文法及び／又は音声理解（speech understanding）及び対話のための規則を含んでもよい。

【００１９】第１の実施例においては、Ｌ１とＬ２は、同じ語彙要素すなわち共通の単語を含んでいない。したがって、$L1 \cap L2 = \{\}$であり、インターフェイス記述のためにマージされた受理言語Ｌは、$L = L1 \cup L2$となる。すなわち、音声装置１内の汎用文書は、第１の機器の言語Ｌ１により記述された機器文書１と、第２の機器の言語Ｌ２により記述された機器文書２から得られる語彙要素と対応するコマンドの対を併合することにより構築される。Ｌ１とＬ２は同じ語彙要素を含まないので、語彙要素は対応するコマンドとともにその語彙要素がどの機器に対してのものであるかを暗示するため、ユーザネットワークコマンドは、適切に生成されて、対応する機器に正しく送信される。

【００２０】第１の実施例において、２つのネットワー

ク機器は、例えばテレビジョンセットとＣＤプレーヤである。このとき、テレビジョンセットに対応するＬ１とＣＤプレーヤに対応するＬ２はそれぞれ、機器文書内で以下の語彙要素からなる。

【００２１】Ｌ１＝｛ＭＴＶ，ＣＮＮ｝

Ｌ２＝｛再生，停止｝

Ｌ１とＬ２は、同じ語彙要素を含まない、すなわちＬ１∩Ｌ２＝｛｝であるので、インターフェイス記述のマージされた受理言語Ｌは、Ｌ＝Ｌ１∪Ｌ２＝｛ＭＴＶ，ＣＮＮ，再生，停止｝となる。例えば、これらの語彙要素はそれぞれ以下のリストに提示されるような機能を有するユーザネットワークコマンドと対応している。

【００２２】

ＭＴＶ→＜テレビジョンをＭＴＶに切り換える＞

ＣＮＮ→＜テレビジョンをＣＮＮに切り換える＞

再生→＜ＣＤプレーヤを再生モードにする＞

停止→＜ＣＤプレーヤを停止モードにする＞

しかしながら、２つの機器が同じ語彙要素すなわち同じ語を共有する場合、Ｌ１∩Ｌ２≠｛｝となる。本発明によれば、これら共通の語彙も識別可能である。本発明の第１の実施例において、機器の名称が各受理言語間、すなわちＬ１とＬ２間の少なくとも共通部分を形成する同じ語彙要素に前置又は後置される。したがって、ユーザは、ユーザの発話コマンドを所望の各機器の名称の前に及び／又は後ろに付加しなければならない。上述のように、コマンドが共通のものでない場合は、機器の名称を付加する必要はないが、付加してもよい。

【００２３】以下の第２の実施例においては、機器の名称は各コマンドの前に付加され、インターフェイスの記述を形成する新しい言語Ｌは、２つの言語Ｌ１とＬ２の間で混同されるおそれのない単語の併合と機器の名称が前置された機器の言語で示される。機器の名称をそれぞれｎ１及びｎ２とすると、

Ｌ＝Ｌ１＼（Ｌ１∩Ｌ２）∪Ｌ２＼（Ｌ１∩Ｌ２）∪ｎ１Ｌ１∪ｎ２Ｌ２

となる。

【００２４】以下では、ＣＤ再生装置及びテープ再生装置を備えるネットワークを例に、上述の実施例を説明し、本発明によるネットワーク機器の制御方法を明らかにする。ここで、ＣＤ再生装置及びテープ再生装置はそれぞれ、「ＣＤ」、「テープ」と命名されている。ＣＤ再生装置の受理言語Ｌ１とテープ再生装置の受理言語Ｌ２をそれぞれ構成する語彙要素を以下に示す。

【００２５】Ｌ１＝｛再生，停止，スキップ｝

Ｌ２＝｛再生，録音，停止，巻き戻し｝

音声インターフェイスの受理言語は、Ｌ＝（Ｌ１＼（Ｌ１∩Ｌ２）∪（Ｌ２＼（Ｌ１∩Ｌ２）∪ｎ１Ｌ１∪ｎ２Ｌ２＝｛スキップ，録音，巻き戻し，ＣＤ再生，ＣＤ停止，ＣＤスキップ，テープ録音，テープ停止，テープ巻き戻し｝となる。対応するユーザネットワークコマンドの機能は、以下のようになる。

【００２６】

スキップ→＜ＣＤの次のトラックへスキップする＞

録音→＜テープ再生装置を録音モードにする＞

巻き戻し→＜テープを巻き戻す＞

ＣＤ再生→＜ＣＤ再生装置を再生モードにする＞

ＣＤ停止→＜ＣＤ再生装置の再生を停止する＞

ＣＤスキップ→＜ＣＤの次のトラックへスキップする＞

テープ再生→＜テープ再生装置を再生モードにする＞

テープ録音→＜テープ再生装置を録音モードにする＞

テープ停止→＜テープ再生装置の再生／録音を停止する＞

テープ巻き戻し→＜テープを巻き戻す＞

【００２７】この第２の実施例では、同一単語問題（same words problem）は、認識されたコマンドが多義的であるとき、ユーザに対して自動的に問い合わせを行い、ユーザに所望する機器を明示させることによって解決される。この場合のシナリオは、形式的には、第１の実施例における識別性を有する言語と同じであるが、その解釈が変わる。この解釈のバリエーションを明示するこの実施例を明確にする第３の実施例は、第２の実施例に基づいており、すなわち、ネットワーク機器としてのＣＤ再生装置及びテープ再生装置に付与されたシナリオに基づいている。受理言語Ｌは、受理言語Ｌ１と受理言語Ｌ２の共有部分に含まれる語彙要素を認識するときに選択処理が行われるという点では、第１の実施例の受理言語Ｌと同様に形成される。すなわち、Ｌ＝Ｌ１∪Ｌ２である。第２の実施例のように、受理言語Ｌ１と受理言語Ｌ２がそれぞれ同じ語彙要素を有するという条件のもとで、音声装置の受理言語Ｌを構成する語彙要素は、Ｌ＝｛再生，停止，スキップ，録音，巻き戻し｝となる。この場合の対応するユーザネットワークコマンドは、

再生→＜明確にせよ＃１＞

停止→＜明確にせよ＃２＞

スキップ→＜ＣＤの次のトラックへスキップする＞

録音→＜テープ再生装置を録音モードにする＞

巻き戻し→＜テープを巻き戻す＞

である。ここで、「明確にせよ＃１」は、音声認識部がどの機器に対するコマンドか（"on what device"）という問い合わせを出力して、受理言語Ｌを受理言語Ｌａに切り換えるということを意味する。Ｌａは、Ｌａ＝｛ＣＤ，テープ｝という語彙要素からなり、対応するユーザネットワークコマンドの機能は、ＣＤ→＜ＣＤ再生装置を再生モードにする＞及び＜Ｌにスイッチバックする＞となり、テープ→＜テープ再生装置を再生モードにする＞及び＜Ｌにスイッチバックする＞となる。

【００２８】またここで、「明確にせよ＃２」は、音声装置１がどの機器に対するコマンドかという問い合わせを出力して、受理言語Ｌを受理言語Ｌｂに切り換えるということを意味する。Ｌｂは、Ｌｂ＝Ｌａ＝｛ＣＤ，テ

ープ｝と示される語彙要素からなり、対応するユーザネットワークコマンドの機能は、ＣＤ→＜ＣＤ再生装置を停止モードにする＞及び＜Ｌにスイッチバックする＞となり、Ｔａｐｅ→＜テープ再生装置を停止モードにする＞及び＜Ｌにスイッチバックする＞となる。

【００２９】本発明の第３の実施例において、第２の実施例に示される音声装置の制御処理を短縮することもできる。この場合、音声装置１は直前に用いられた機器をデフォルト機器とみなして、音声装置１とデフォルト機器の統合された受理言語Ｌからなる汎用文書に基づいて、受信され認識されたユーザの発話コマンドを対応するユーザネットワークコマンドに変換する。直前に用いられた機器以外の他の機器を再度制御する場合にも、音声装置の制御処理を短縮することができる。この場合、例えば単に機器の名称を発することによって、機器を制御するよう音声装置に指示する。

【００３０】第４の実施例では、ダイアログの適応化が行われる。上述のような状況においては、多義的な語を正しく割り当てるために、機器明示の問い合わせが、ユーザに対して出力され、システムはその問い合わせに対するユーザの反応を認識する。機器明示のための問い合わせに対するユーザの回答の大部分が直前に指定された機器である場合、問い合わせをスキップし、ユーザネットワークコマンドを直前の機器に直接供給するようにしてもよい。また、コマンドがどの機器に対して出されたのかを予測するために、複数の情報源を用いてもよい。例えば、直前に使用された機器を考慮するだけでなく、各機器の類似点又は機器の分類も考慮する。例えば、記録コマンドは、ユーザがテレビを観ている間に出される場合は、オーディオテープへの記録ではなく、ビデオテープへの記録である可能性が高い。一方、ユーザがラジオを聴いている間に出される記録コマンドは、オーディオテープへの記録である可能性が高い。そこで、使用される可能性の高いネットワーク機器を、例えば電源が投入されているネットワーク機器から選択するようにしてもよい。また、機器の分類は、ユーザの行為又は機器の機能に関する記述を調査することによって学習することもできる。

【００３１】上述の全ての実施例において、ネットワーク機器がネットワークに接続された後又は、機器がユーザネットワークコマンドを音声装置１から受信した後に、各ネットワーク機器は、機器文書を音声装置１に直接送信することができる。また、現在のステータスから独立した機器文書等を音声装置に送信してもよい。機器は、複数の機器文書を有していてもよい。または、機器は、内容情報（contentinformation）の変更に基づいて動的に機能を変えて、その変化を認めるときにダイアログ及び音声認識能力を更新してもよい。あるいは、新しい文書を音声装置に送信することによって、定期的に機能を変えてもよい。機器文書は、対応する機器内に格納されずに、例えば装置製造者のインターネットサーバのような機器文書提供装置内に格納されてもよい。この場合、音声装置１は上記の機器がネットワークに接続されていると認識すると、機器文書をダウンロードしてもよい。

【００３２】機器は、それぞれが異なる国の言語による機器全体の受理言語を含む複数の文書を有することができる。例えば、このような機器は、最初にドイツ語の機器文書を音声装置に送信する。続いて、ユーザから対応するコマンドを受信した後に、英語の機器文書を音声装置に送信することができる。これにより、音声装置は、ドイツ語又は英語のいずれかのユーザコマンドをユーザネットワークコマンドに変換して機器を制御することができる。

【００３３】語彙要素の他に、認識又は合成を行うときに用いられる語のための発音の定義を機器文書の要素として含ませてもよい。そして、それらの発音の定義を音声装置の汎用文書に直接統合してもよい。発音の定義が言語に含まれるとき、それらの発音の定義は、例えば、１つの機器用の認識及び合成部に専用に用いてもよく、又は、全ての機器で共有してもよい。すなわち、第１の機器によって導入された発音の定義は、第２の機器、すなわちネットワークに接続された他の機器に対して発せられた語に対しても有効とすることができる。

【００３４】合成を行うために、複数の機器文書を統合し、音声装置内で、汎用文書の適応化を行ってもよい。問い合わせに対する回答を１つの機器だけでなく、複数の機器に対して行ってもよい。複数の機器からの発話の合成は、同時に出力されないことを保証する必要がある。発話の中には、供給優先度に基づいて、例えば警告のような外部の事象によって生じるものがある。例えばそれらの発話は、他の発話に割り込み、ユーザの非活動期間が終わると、割り込まれた発話は再導入され、ダイアログは続行する。ユーザの入力及びユーザの発話に一致する尤度が高いと考えられるストリームが継続している間、全てのストリームがアクティブな状態にあるという事実を前提として、複数のダイアログのストリームが平行して管理される。

【００３５】音声装置の汎用文書は、初期的には空であってもよく、あるいは、１又は複数のネットワーク機器のための、インターフェイスを記述する要素の基本セットを有していてもよい。後者の場合、機器から音声装置に送信される文書は、具体性を有さず、汎用文書を部分的に参照するものであってもよい。

【００３６】例えば、特定の機器のためのダイアログ文法のみを機器文書において特定し、発話の言い回しのためのデフォルト文法を汎用文書のユーザコマンド解釈要素の基本的なセットに含ませてもよい。また、音声認識部の汎用文書内に、ある機能を完全に記述してもよい。例えば、電子乗組ガイドを、音声ダイアログによって

どのように制御するのかを記述する。この場合、デフォルト文法及び一連の単語のための文法には、例えば、番組名、製作者名、俳優名、時間帯及び放送日に関する情報を示す変数が含まれており、電子番組ガイド装置から音声装置に送信される文書は、これら変数を満たす情報のみを含む。

【００３７】他の実施例においては、機器から音声装置に送信される機器文書は、ネットワークコマンドに対応するキーワードとカテゴリ識別子とを含む。音声装置は、これらのカテゴリ識別子に基づいて、発生し得る連続するユーザの発話に対して、単語列に関するどの文法を適用するかを決定する。例えば、キーワードが「ボイス・オブ・アメリカ（voice of America）」であり、カテゴリが「ラジオ」である場合、ユーザは、「ボイス・オブ・アメリカを聴きたい（I want to listen to voice of America）」又は「ボイス・オブ・アメリカをオン（Turn on voiceof America）」等と発生する。また、カテゴリが「ラジオカセットレコーダ」である場合、発生し得るユーザの発話は、「ボイス・オブ・アメリカを録音（Please record voice of America）」等である。

【００３８】汎用文書の適応化を行うために、音声理解の基礎を形成する概念／値の組、例えばある処理を開始するために必要な（時刻，８：００）といった組を音声装置に送信される機器文書に機器用の文法として含ませてもよい。音声インターフェイスを記述する機器文書は、規則を含み、この規則は、単語列と概念との間のマッピング及びどのように変数値を埋めるかを定義する。さらに、機器文書は、概念／値を一対の組から動作へのマッピングを定義する規則を有する。例えば（コマンド，記録）、（タイトル，風とともに去りぬ）及び（時刻，８：００）という組は、ＶＴＲの録画処理を表す。ダイアログは、ユーザ側からはじめてもよく、すなわち発話の順番は、ユーザを主導としてもよい。発話の順番とは、例えば、「８時、映画、風とともに去りぬ」等である。１又は複数の発話で情報が伝えられる場合は、発話の順番は、ユーザに任せられる。システムは、例えば、文法に基づきそして各ネットワーク機器に送信されたユーザネットワークコマンドによって動作を始動するために必要な情報を収集する。このシステムは、例えば、付加的な文法を含むさらなる文書を音声装置に送信し、音声装置は、情報が欠落している場合、ユーザに情報を聞き返すことができ、また、新しく認識されたユーザの発話をさらなるユーザネットワークコマンドに割り当てることもできる。

【００３９】機器文書は、ユーザコマンド解釈要素として、語彙要素、発音、単語列のための文法、音声理解及びダイアログのための規則に関する情報の他に、同一又は同等のユーザネットワークコマンドに割り当てられる複数言語に対応する同様な情報、すなわち語彙要素、発音、単語列のための文法、音声理解及びダイアログのための規則関する情報を含んでもよい。この情報によって、ユーザはいかなる言語によっても機器を制御することができる。上述のように、この情報を同じ機器が有する複数の機器文書中に含ませてもよい。これによって、複数言語の音声による処理を実現することができる１つの音声装置のインターフェイスを特定できる。上記の概念／値の対は言語独立であるので、ダイアログの文法は言語独立とすることができる。

【００４０】上述のように、機器文書は、発話のための文法によって明示的にあるいは暗示的に付与された語彙を含む。１又は複数の発音を各語彙要素、すなわち各語に対して任意に付与することもできる。任意の発音が機器文書に付与されないときは、これを自動的に生成してもよい。しかしこの場合、特に単語が固有名詞、外来語、略語である場合や方言又は外国語による特殊な発音を認識しなくてはならない場合、エラーが発生しやすい。発音は、単音又は異音のモデルのセットが与えられた認識部内に単語のモデルを生成するための基礎となる。

【００４１】機器文書は、複数の利用可能な機器に関する複合コマンドを作成するユーザコマンド解釈要素及び関連するユーザコマンドの他に、標準化された情報、例えば「ＶＴＲ」のような機器のカテゴリに関する情報及び／又は「記録可能なテープを有する」等の機能に関する情報を含んでもよい。例えばユーザは、テレビジョンセットに表示される電子番組ガイドから将来放映される予定の映画を選択するとき、「この映画を録画して下さい（Please record this movie）」と発する。つまり、ネットワークコマンドが、適切なチャンネル、日付及び時間帯に関する情報を番組ガイド装置からＶＴＲに送信する必要がある。そして、適切な録画が行われることを保証するコマンドをＶＴＲに送信する必要がある。これに類似した状況として、映画をあるＶＴＲテープから別のＶＴＲテープにダビングする場合がある。このような状況では、各機器の音声インターフェイス記述は、通常、このような発話、すなわち適切なチャンネル、日付及び時間帯等に関する発話を定義しない。したがって、音声装置は、推論部による推論に基づいて汎用文書を適応化してもよい。推論部は、推論構成要素として機器文書内に含まれており、単語列に関する文法及びダイアログを含む音声インターフェイスの記述を複数の機器に関する複合コマンドの１つに含まれる可能性のある追加的機能に提供する。

【００４２】ネットワーク機器を制御するための全ての情報は、音声装置内の１つの汎用文書に含まれるので、処理あるいは分析を容易に行うことができる。特に、ユーザコマンド解釈要素の異なる利用可能な内容間又は／及び汎用文書内の同様又は同等のユーザネットワークコマンドに割り当てられた二重の情報間に関連があるときは特に容易である。

【００４３】したがって、本発明によると、ネットワークが備える音声装置を介して制御することのできるネットワーク機器は、機器の機能及び音声インターフェイスを記述する機器文書を音声装置に送信することができる。音声装置は、機器文書を汎用文書に統合する。汎用文書は、接続されたネットワーク機器を制御するために、認識されたユーザコマンドをユーザコマンド解釈要素を用いてユーザネットワークコマンドに変換するための基礎を形成する。機器文書は例えばユーザコマンド解釈要素及び関連するユーザネットワークコマンドのような語彙からなる。さらに、複数言語に対応するこれらと同様な情報、又は、音声理解における動的なダイアログに関する情報だけでなく、発音、単語列のための文法、音声理解及びダイアログのための規則を機器文書のユーザコマンド解釈要素内に含めてもよい。１つの機器が複数の機器文書を有し、必要時にそれらを動的に音声装置に送信してもよい。さらに、ネットワーク機器は、音声機能に関する仕様を実行時に音声装置に送信するので、本発明を適用することにより、機器は、機器文書の内容の変更に基づいてその機能を動的に変えることができる。機器がその機能を動的に変える場合、ネットワーク機器は、動的に機器文書を生成する、又は、既存の機器文書を更新する。更新は、例えば、放送局の名称を更新する又は挿入することによって行われ、ネットワーク機器は、更新された機器文書を音声装置に送信する。

【００４４】

【発明の効果】以上のように、本発明に係るネットワーク機器の制御方法では、ネットワーク機器に対応し、ユーザコマンド解釈要素と関連するユーザネットワークコマンドの少なくとも１つの対からなる言語を含む少なくとも１つの機器文書を受信し、受信された機器文書を、機器文書の言語と同様に構成される音声装置の言語からなる汎用文書に適応化する適応化し、受信及び認識されたユーザの音声コマンドを、ユーザコマンド解釈要素と汎用言語に含まれる関連するユーザネットワークコマンドの全ての対に基づいて、対応するユーザネットワークコマンドに変換する。これにより、本発明に係るネットワーク機器の制御方法によれば、音声装置を備えるネットワークに接続された複数のネットワーク機器を制御するユーザネットワークコマンドを簡易且つ効率的に管理、処理、及び分析することができる。

【図面の簡単な説明】

【図１】本発明を適用した音声装置の構成を示すブロック図である。

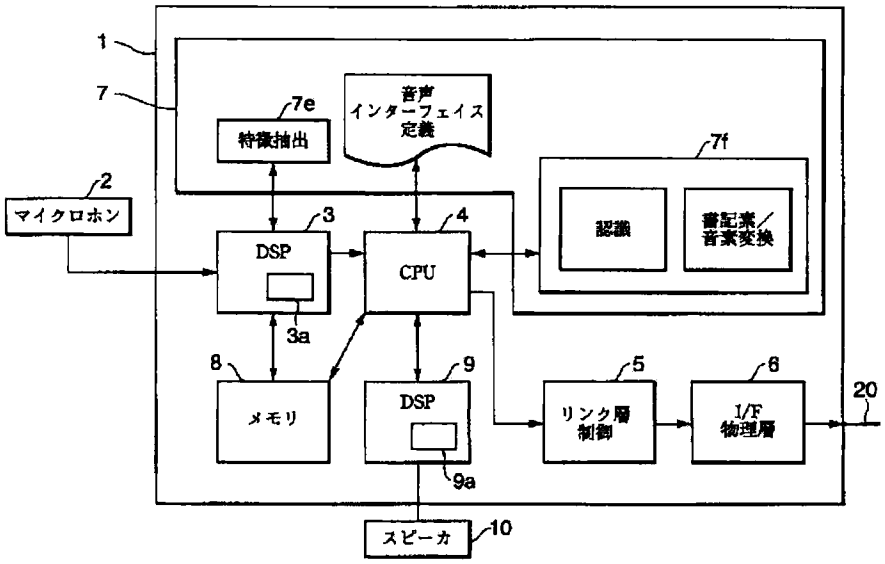
【図２】本発明を適用したネットワーク機器の構成を示すブロック図である。

【図３】従来の音声装置の構成を示すブロック図である。

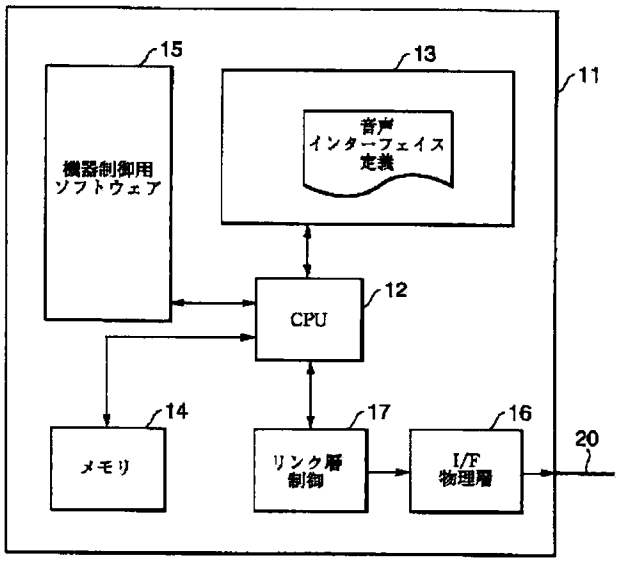
【符号の説明】

１　音声装置、２　マイクロホン、３　ＤＳＰ、４　ＣＰＵ、５　リンク層制御部、６　Ｉ／Ｆ物理層部、７　メモリ、８　メモリ、９　ＤＳＰ、１０スピーカ、２０　バス

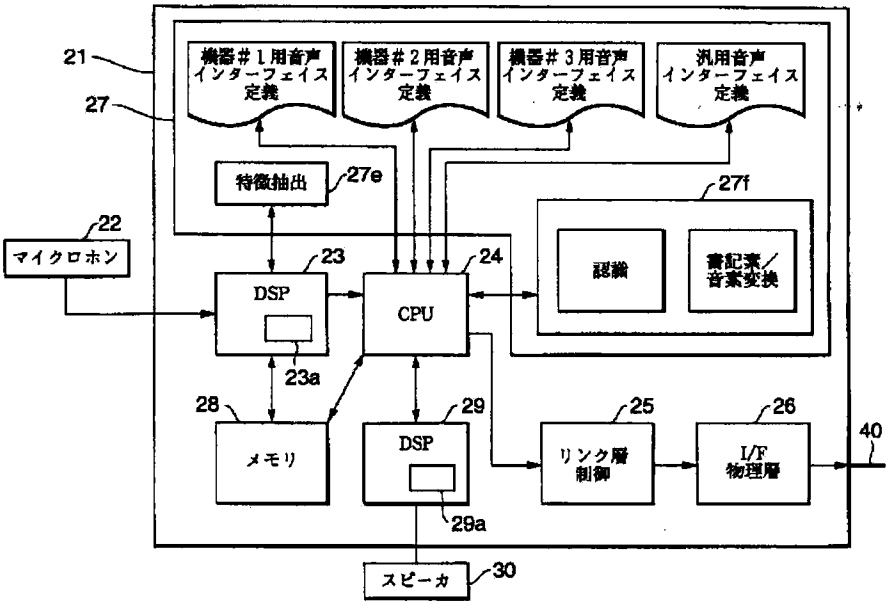
【図１】



【図２】



【図３】



フロントページの続き

(51) Int.Cl.[7]	識別記号	F I	ﾃｰﾏｺｰﾄﾞ (参考)
H04Q 9/00	331	G10L 3/00	561H

(72)発明者　ステファン　ラップ
ドイツ連邦共和国　ディーー70736　フェルバッハシュトゥットゥガルター　シュトラーセ　106　ソニー　インターナショナル（ヨーロッパ）　ゲゼルシャフト　ミット　ベシュレンクテル　ハフツング　シュトゥットゥガルト　テクノロジーセンター内
(72)発明者　シルケ　ゴロンジー
ドイツ連邦共和国　ディーー70736　フェルバッハシュトゥットゥガルター　シュトラーセ　106　ソニー　インターナショナル（ヨーロッパ）　ゲゼルシャフト　ミット　ベシュレンクテル　ハフツング　シュトゥットゥガルト　テクノロジーセンター内
(72)発明者　ラルフ　コンペ
ドイツ連邦共和国　ディーー70736　フェルバッハシュトゥットゥガルター　シュトラーセ　106　ソニー　インターナショナル（ヨーロッパ）　ゲゼルシャフト　ミット　ベシュレンクテル　ハフツング　シュトゥットゥガルト　テクノロジーセンター内
(72)発明者　ペーター　ブフナー
ドイツ連邦共和国　ディーー70736　フェルバッハシュトゥットゥガルター　シュトラーセ　106　ソニー　インターナショナル（ヨーロッパ）　ゲゼルシャフト　ミット　ベシュレンクテル　ハフツング　シュトゥットゥガルト　テクノロジーセンター内
(72)発明者　フランク　ジロン
ドイツ連邦共和国　ディーー70736　フェルバッハシュトゥットゥガルター　シュトラーセ　106　ソニー　インターナショナル（ヨーロッパ）　ゲゼルシャフト　ミット　ベシュレンクテル　ハフツング　シュトゥットゥガルト　テクノロジーセンター内
(72)発明者　ヘルムート　ルッケ
ドイツ連邦共和国　ディーー70736　フェルバッハシュトゥットゥガルター　シュトラーセ　106　ソニー　インターナショナル（ヨーロッパ）　ゲゼルシャフト　ミット　ベシュレンクテル　ハフツング　シュトゥットゥガルト　テクノロジーセンター内